

CLAIMS

What is claimed is:

1. A method for providing content to a content repository, comprising:
 providing a process operable to interact with a virtual content repository (VCR) and capable of communicating with the VCR using a computer network;
 providing a mechanism for the process to interact with the VCR;
 identifying a first content;
 associating a first schema with the first content;
 providing to the VCR at least one of: 1) the first content; 2) a reference to the first content; and 3) the first schema to the VCR; and
 wherein the VCR is operable to provide to the at least one content repository the at least one of: 1) the first content; 2) the reference to the first content; and/or 3) the first schema.
2. The method of claim 1 wherein:
 the mechanism for interacting with the VCR includes an Application Programming Interface (API).
3. The method of claim 1 wherein:
 the VCR integrates the at least one content repository into a logical content repository.
4. The method of claim 1 wherein:
 each one of the at least one content repositories exposes a first set of services to enable its integration into the VCR.
5. The method of claim 1 wherein the step of identifying the first content includes:
 traversing a file system and/or a website.
6. The method of claim 1 wherein the step of identifying the first content includes:
 extracting properties from one of: 1) a file; 2) a hypertext markup language

(HTML) document; and 3) an Extensible Markup Language (XML) document.

7. The method of claim 1 wherein the step of associating the first schema with the first content includes:

acquiring the first schema from at least one of: 1) a file; 2) a hypertext markup language (HTML) document; and 3) an Extensible Markup Language (XML) document.

8. The method of claim 1 wherein the step of providing the first content and/or the first schema to the VCR includes:

persisting in the at least one content repository the at least one of: 1) the first content; 2) the reference to the first content; and/or 3) the first schema.

9. The method of claim 1 wherein the step of providing the first content and/or the first schema to the VCR includes:

preserving in one of the at least one content repositories hierarchical relationships between the first content and other content in the VCR.

10. A method for providing content to a content repository, comprising:

providing a process operable to interact with a virtual content repository (VCR) and capable of communicating with the VCR using a computer network;

providing a mechanism for the process to interact with the VCR;

identifying a first content;

associating a first schema with the first content;

providing at least one of the following to the VCR: 1) the first content; 2) a reference to the first content; and 3) the first schema to the VCR; and

wherein the VCR integrates at least one content repository into a logical content repository.

11. The method of claim 10 wherein:

the mechanism for interacting with the VCR includes an Application Programming Interface (API).

12. The method of claim 10 wherein:

the VCR is operable to provide to the at least one content repository the at least one of: 1) the first content; 2) the reference to the first content; and/or 3) the first schema.

13. The method of claim 10 wherein:

each one of the at least one content repositories exposes a first set of services to enable its integration into the VCR.

14. The method of claim 10 wherein the step of identifying the first content includes:

traversing a file system and/or a website.

15. The method of claim 10 wherein the step of identifying the first content includes:

extracting properties from one of: 1) a file; 2) a hypertext markup language (HTML) document; and 3) an Extensible Markup Language (XML) document.

16. The method of claim 10 wherein the step of associating the first schema with the first content includes:

acquiring the first schema from at least one of: 1) a file; 2) a hypertext markup language (HTML) document; and 3) an Extensible Markup Language (XML) document.

17. The method of claim 10 wherein the step of providing the first content and/or the first schema to the VCR includes:

persisting in one of the at least one content repositories the at least one of: 1) the first content; 2) the reference to the first content; and/or 3) the first schema.

18. The method of claim 10 wherein the step of providing the first content and/or the first schema to the VCR includes:

preserving in one of the at least one content repositories hierarchical relationships between the first content and other content in the VCR.

19. A content mining system for providing content to at least one content

repository, comprising:

- a first process operable to interact with a Virtual Content Repository (VCR);
- a first set of services operable to enable integration of the at least one content repository into the VCR;
- a second set of services operable to enable interaction between the first process and the VCR;
- wherein the first process is operable to provide to the VCR at least one of: 1) content; 2) a reference to the content; and 3) a schema corresponding to the content; and
- wherein the VCR is operable to integrate the at least one content repository into a logical repository.

20. The system of claim 19, further comprising:

- at least one second process operable to interact with the first process;
- wherein the at least one second process is operable to provide to the first process the at least one of: 1) content; 2) a reference to the content; and 3) a schema corresponding to the content; and
- a third set of services operable to enable interaction between the at least one second process and the first process.

21. The system of claim 20 wherein:

- the third set of services provides a first function for directing the at least one second process to extract at least one property from the content; and
- wherein a property is an association between a name and a value.

22. The system of claim 20 wherein:

- the at least one second process can derive the schema from the content.

23. The system of claim 19 wherein:

- the content can include at least one property; and
- wherein a property is an association between a name and a value.

24. The system of claim 19, further comprising:

- at least one second process operable to derive the at least one property from

the content.

25. The system of claim 19, further comprising:
at least one second process operable to locate the schema corresponding to the content.
26. The system of claim 19, further comprising:
at least one second process operable to extract the content and/or the schema from at least one of: 1) a file; 2) a hypertext markup language (HTML) document; and 3) an Extensible Markup Language (XML) document.
27. The system of claim 19 wherein:
the first process is operable to recursively traverse a file system and/or a website.
28. The system of claim 19 wherein:
the first set of services and the second set of services share a content model.
29. A system, comprising:
means for providing a process operable to interact with a virtual content repository (VCR) and capable of communicating with the VCR using a computer network;
means for providing a mechanism for the process to interact with the VCR;
means for identifying a first content;
means for associating a first schema with the first content;
means for providing at least one of the following to the VCR: 1) the first content; 2) a reference to the first content; and 3) the first schema to the VCR; and
wherein the VCR is operable to provide to the at least one content repository at least one of: 1) the first content; 2) a reference to the first content; and 3) the first schema to the VCR.
30. A computer data signal embodied in a transmission medium, comprising:
a code segment including instructions to provide a process operable to interact with a virtual content repository (VCR) and capable of communicating with the VCR

using a computer network;

a code segment including instructions to provide a mechanism for the process to interact with the VCR;

a code segment including instructions to identify a first content;

a code segment including instructions to associate a first schema with the first content;

a code segment including instructions to provide to the VCR at least one of:

1) the first content; 2) a reference to the first content; and 3) the first schema to the VCR; and

wherein the VCR is operable to provide to the at least one content repository the at least one of: 1) the first content; 2) the reference to the first content; and/or 3) the first schema.

31. A machine readable medium having instructions stored thereon that when executed by a processor cause a system to:

provide a process operable to interact with a virtual content repository (VCR) and capable of communicating with the VCR using a computer network;

provide a mechanism for the process to interact with the VCR;

identify a first content;

associate a first schema with the first content;

provide to the VCR at least one of: 1) the first content; 2) a reference to the first content; and 3) the first schema to the VCR; and

wherein the VCR is operable to provide to the at least one content repository the at least one of: 1) the first content; 2) the reference to the first content; and/or 3) the first schema.

32. The machine readable medium of claim 31 wherein:

the mechanism for interacting with the VCR includes an Application Programming Interface (API).

33. The machine readable medium of claim 31 wherein:

the VCR integrates the at least one content repository into a logical content repository.

34. The machine readable medium of claim 31 wherein:
each one of the at least one content repositories exposes a first set of services to enable its integration into the VCR.
35. The machine readable medium of claim 31, further comprising instructions that when executed cause the system to:
travers a file system and/or a website.
36. The machine readable medium of claim 31, further comprising instructions that when executed cause the system to:
extract properties from one of: 1) a file; 2) a hypertext markup language (HTML) document; and 3) an Extensible Markup Language (XML) document.
37. The machine readable medium of claim 31, further comprising instructions that when executed cause the system to:
acquire the first schema from at least one of: 1) a file; 2) a hypertext markup language (HTML) document; and 3) an Extensible Markup Language (XML) document.
38. The machine readable medium of claim 31, further comprising instructions that when executed cause the system to:
persist in one of the at least one content repositories the at least one of: 1) the first content; 2) a reference to the first content; and 3) the first schema to the VCR.
39. The machine readable medium of claim 31, further comprising instructions that when executed cause the system to:
preserve in one of the at least one content repositories hierarchical relationships between the first content and other content in the VCR.